封面页

# 摘要

社会经济不断增长，人们的生活水平逐渐提高，不断促进着监控技术和网络技术的革新，高清网络监控摄像头也被广泛应用在各个场所中，以满足社会各方面的安防需求。如今网络远程监控已经被普及，摄像头数量也日益增长，伴随而来的是海量监控视频，但是却无法找到足够的人力资源和时间资源对这海量视频数据进行处理。将视频监控智能化、自动化成为当前监控领域的迫切需求。

本文针对上述海量监控视频的问题，开发基于视频摘要的视频内容检索系统，给出了一种基快速的生成视频摘要方法，同时支持实时监控视频和离线监控视频处理。视频摘要是监控视频的浓缩，提取视频中有意义的部分，然后浓缩成一个很短的摘要，让监控人员能用最短的时间看完一个长的监控视频。视频内容检索是在提取视频摘要的基础上，对浓缩的监控视频的事件进行简单的特征搜索，避免了对整个视频的分析，大大缩小了视频内容检索时间。

视频摘要算法包括运动物体检测、运动物体跟踪以及摘要合成三个部分，先用帧间差分法检测出运动物体的矩形轮廓，然后用矩形轮廓匹配对每个运动物体进行轨迹跟踪，最后遍历每一个运动事件生成视频摘要。视频内容检索主要是对一些简单特征进行检索，如事件发生时间，对象的运动方向，入侵区域，对象的颜色等。

本文先对基于视频摘要的视频内容检索系统进行功能介绍，描述其适用场景范围和存在的问题；然后对系统整体流程和各部分算法进行详细介绍；最后对系统的实验结果进行分析，提出不足之处和可以改进的地方。

**关键字：视频摘要，视频内容检索，图像特征提取，运动物体检测与跟踪**

# Abstract

# 目录

1. **绪论**
   1. **背景与意义**

随着社会的迅速进步和科学技术的发展，安防监控在软硬件层面上都在不断地革新，安防监控摄像头也被广泛应用于社会中的每个角落。现在低分辨率低码率的摄像头已经不能满足监控的需求，高清监控摄像头是安防监控的必然趋势，高清监控视频对于视频取证，人物识别，交通监控等有着重大意义。由于监控系统己经基本覆盖社会的各个角落，其所产生的海量视频数据无法有效进行有效的管理和查看。实际的监控任务仍需要较多的人工完成，而且现有的视频监控系统通常只是录制视频图像，也有一些厂家提供了一些自动监控特定事件的软件，但实际应用较少。一般的监控视频只能用作事后取证，而且人工浏览监控视频进行取证需要花费很长时间。

随着高清监控摄像头的不断普及，监控视频将在取证和特定目标识别等方面发挥重大作用。面对海量的监控视频，智能监控也是必然趋势，将智能监控应用到实时高清摄像头中成为了当务之急。

针对上述提到的问题，我提出了对应的解决方案，即基于摘要的监控视频信息检索系统,该系统先在实时监控阶段对实时视频流进行摘要分析（摘要分析也可以对离线视频进行处理），然后在事后取证查找过程再加入自动检索功能。视频摘要是视频内容的一个浓缩，去除无用部分，只保留有意义的部分。先进行视频摘要处理，可以为视频检索过程节省大量时间。当需要浏览监控视频时可以只浏览视频摘要，或者根据对象的特征进行搜索，快速找到监控视频内容。

视频摘要的目的是提取视频中有意义的内容（即运动事件），因此此系统不适用于人来人往的闹市，或者有很多目标固定存在的场所（比如说大型办公室等）。本系统可以应用于出入口（比如说公司大门）监控，小区道路监控，非塞车时的交通道路监控，别墅监控等。

* 1. **国内外相关产品概况**
     1. **BriefCam**

BriefCam是第一个成功商业应用的视频摘要软件，由耶路撒冷希伯来大学凭借其技术实力研发。BriefCam实现了以前全球对于视频查看未能满足的需求，即快速地调查和查明事件并采取行动。BriefCam视频摘要技术使用户可以用一分钟的时间回顾24小时内发生的事件，将数小时的事件压缩成“摘要”，仅用几分钟来查看。其客户包括警察、军队、边检及其他国土安全机构，以及政府安全部门、机场、铁路、海运及其他运输部门、银行、办公楼管理、零售商及更多。

* + 1. **华尊视频摘要**

华尊视频摘要分析技术，采用基于视频对象的技术原理。追踪和分析视频中的活动对象，并提取对象的运动区域、颜色、大小等信息，从而建立与原始视频的索引关系。通过构建视频摘要剪辑，显示完整的视频内容，可将24小时内发生的所有事件以浓缩短片的形式，在短短几分钟内完整显示出来。华尊视频摘要最大的特点就是增加了一些对象特征提取的功能。

* + 1. **QBIC系统**

QBIC（Query By Image Content）由IBM Almaden研究中心所开发，是“基于内容”视频检索的典型代表。此检索系统利用形状、纹理、颜色和对象运动等特征来描述视频内容，并在此基础上实现视频内容检索。QBIC同时提供了基于内容的视频信息检索手段和静止图像信息检索手段，用户可以使用颜色特征、纹理特征、例子图像、草图和对象运动等信息对视频或者图像进行检索。在视频内容分析方面运用了运动估计、代表帧生成、镜头检测、层描述等多种分析技术。

* + 1. **Visual Seek系统**

Visual Seek是美国哥伦比亚大学实现的在互联网上使用的基于内容的检索系统。它实现了互联网上的基于内容的视频或者图像检索系统，提供了一个供用户能在互联网上搜索和检索视频以及图像的工具。

* 1. **智能监控系统的市场预测和发展趋势**
     1. **智能监控技术不断革新**

从软硬件技术方面来看，首先是半导体工艺的发展，根据摩尔定律，处理器性能逐年升高，即是是计算量很大的算法也能很快的处理完；其次是视频图像处理技术也在不断的完善和革新，能够处理更为复杂和多变的场景，能够分析和识别更多的异常事件和行为，处理的时间复杂度也不断的优化降低；再次是监控设备也不断得到更新，高清视频监控也已经得到很好的应用。

* + 1. **不断发展的市场**

其实，对智能监控系统的需求一直存在，只是智能监控技术还不够成熟，目前智能视频监控的应用场景还比较有限，主要应用于一些特定的场合，但随着软硬件技术的不断进步，智能视频监控必将覆盖到每一个监控摄像头，甚至走进千家万户。

根据现在的发展趋势，安防监控将会在未来几年内基本实现城市高清监控，高清视频使得智能监控技术发挥更多的用武之地，之前由于视频分辨率太低图像质量太差等原因而导致无法分析的情况不复存在。

总之，智能安防监控的市场需求在不断扩大，应用领域也越来越广，甚至可以作为一个模块集成到数字家庭系统中。

* 1. **开发环境和开发工具简介**

本系统使用Qt Creator作为IDE和UI开发工具进行开发，同时用到了Opencv和FFmepg两个开源跨平台的工具。

* + 1. **Qt Creator**

Qt Creator是一个跨平台的IDE，同时支持Windows、Linux和Mac OS X系统，开发人员能用QT更方便快捷的完成开发任务。Qt Creator整合了跨平台的自动化构建系统：qmake与CMake，使用Qt Designer可以很方便快捷的设计和构建图形界面。此外，Qt Creator还具有对C++语言的完整表达式检查，上下文关联，代码不全，键入代码时的行间错误即时指示等功能。

* + 1. **Opencv**

OpenCV的全称是：Open Source Computer Vision Library。它是一个基于BSD许可证授权（开源）发行的跨平台计算机视觉库，可以运行在Linux、Windows和Mac OS操作系统上。它轻量级而且高效地实现了图像处理和计算机视觉方面的很多通用算法。

目前OpenCV广泛运用在人机互动、物体识别、图像分割、人脸识别、动作识别、运动跟踪、机器人等领域。

因为OpenCV中提供了很多方便高效的图像处理算法找到，特别是运动物体的检测与跟踪，故采用OpenCV 进行辅助开发。

* + 1. **FFmpeg**

FFmpeg是一个开源免费跨平台的视频和音频流方案。其包含了非常完善的视频解码技术，同时还支持网络RTSP视频流的解码。使用FFmpeg工具可以很方便解决主流监控视频的解码问题。

* 1. **主要研究内容和论文主要结构**

本论文主要研究视频摘要和视频内容检索，视频摘要又可分为视频中运动物体检测与跟踪，视频摘要生成等；视频内容检索主要是在视频摘要的基础上根据运动对象的颜色、大小、轨迹、时间等特征进行检索。

本论文共分为六章：

第一章主要讨论了基于视频摘要的视频内容检索。主要是描述其背景和意义，对目前现有的相关系统进行简单的介绍。

第二章对基于视频摘要的视频内容检索系统进行需求分析。

第三章主要研究了视频摘要的实现。针对实时监控视频，提出了一种快速生成视频摘要的方法。该方法可以应用于多种监控场景中，事件检出率高，速度快，满足高清网络摄像头的实时监控需求。

第四章主要研究了一些图像特征提取技术，包括颜色、形状、类别等。在利用特征进行视频内容检索的时候，采用逐个特征过滤的方式，从最易提取的特征开始过滤，使检索的效率最大化。

第五章给出了整个系统的架构的实现，然后对系统的运行测试结果进行详细的分析，指出不足之处和改进的方向。

第六章对论文作出了总结和展望。总结了论文的主要研究工作与创新之处，并指出其中的不足；对基于云处理的视频摘要和视频内容检索的发展方向进行了展望。同时描述了本人的下一步工作内容。

1. **系统需求分析**
2. **系统功能概览**

基于摘要的监控视频内容检索系统旨在开发一个软件系统，该系统能对监控视频进行实时或离线的摘要分析，然后根据特征进行视频信息检索。系统主要分为两大功能：视频摘要和视频信息检索。下图是本系统的功能列表：

|  |  |  |
| --- | --- | --- |
| **功能** | **子功能** | **功能点** |
| 本地/实时视频摘要 | 本地视频文件读取 | 支持主流视频格式，包括avi，mp4，wmv，flv，rmvb等 |
| 摄像头视频流读取 | 支持usb摄像头和rtsp视频流读取 |
| 事件跟踪 | 对视频中的运动对象进行检测与跟踪 |
| 事件统计 | 实时统计监控视频中出现的事件数量 |
| 生成视频摘要 | 将不同时间发生的事件整合到同时间的场景中 |
| 生成分析文件 | 生成分析文件，记录监控视频中所有事件的信息 |
| 摘要事件查看 | 视频摘要查看 |
| 回溯原始视频 |
| 视频信息检索 | 根据事件发生时间检索 | 用户指定一个时间区间，检索该时间区间的事件 |
| 根据事件的运动方向检索 | 根据事件中目标的运动方向进行检索 |
| 根据事件的入侵区域检索 | 根据事件中目标的入侵区域进行检索 |
| 根据事件的目标大小检索 | 根据事件中目标的面积大小进行检索 |
| 根据事件颜色特征检索 | 根据事件中目标的主要颜色进行检索 |
| 根据事件的目标类型检索 | 根据目标的类别（分为人、车、物体）进行检索 |

系统功能列表

1. **视频摘要**

视频摘要就是监控视频的浓缩，也可以说成视频分析功能，对监控视频进行运动物体检测与跟踪，将视频中所有有意义的事件（运动事件）提取出来并生成视频的摘要，然后将事件信息保存。这里要求系统支持主流的视频格式，并且支持实时监控视频流分析，能迅速地分析整个视频，并将视频摘要信息提取出来，然后把这个视频的所有摘要（运动事件）信息按照一定格式保存到本地文件，这样做的目的是进行摘要备份，避免二次分析视频导致浪费时间。分析完视频之后生成摘要视频和事件列表，供用户查看。

1. **本地视频文件读取**

【功能描述】监控员从本地选取一个视频文件。

【前提条件】视频文件是主流视频格式，没有损坏。

【业务角色】监控员。

【业务流程】练习者（或者管理员）登录的泳道式业务流程图如下：



对于管理员登录，只是进入高尔夫球挥杆自动分析软件的本地存储界面，可以对教练视频和明星视频进行更新（添加或者删除）。

【输入数据】

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **字段名** | **输入形式** | **必填** | **内容及规则** |
| 1 | 用户名 | 文本框录入 | 是 | 用户名的规则和接口由甲方提供 |
| 2 | 密码 | 密码框录入 | 是 | 密码规则和接口由甲方提供 |
| 3 | 用户类型 | 单选 | 是 | 练球者、管理员，默认：练球者 |

【输出数据】

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **字段名** | **输出形式** | **可修改** | **内容及规则** |
| 1 | 登录成功 | 对话框 | 否 | 格式：欢迎您，张三！ |
| 2 | 登录失败 | 对话框 | 否 | 场景一：对不起，系统中不存在您输入的用户名！  场景二：对不起，您输入的密码与用户名不匹配，请重新输入！ |

【约束和例外处理】

1、用户点击“登录”按钮后，系统要在2秒内做出响应。

是

1. **摄像头视频流读取**
2. **事件检测与跟踪**
3. **事件统计**
4. **生成视频摘要**
5. **生成分析文件**
6. **摘要事件查看**
7. **视频信息检索**

**2.3.1时间特征检索**

从原始视频获取到的图像是彩色图像，如果直接对彩色图像操作，计算量会非常大，为了提高算法处理速度，要先将彩色图像转化为灰度图。

彩色RGB图像的每一个像素点有三个分量，分别代表红、绿、蓝三个通道的值。由RGB颜色空间的图像转化为灰度图最简单的做法是取R、G、B三个分量的平均值，但是这种做法并不科学，因为人眼对红、绿、蓝三种颜色的感知是不一样的，灰度值其实就是亮度，由RGB像素转化为灰度像素的公式为：

在一些处理器中，整数的运算比浮点数运算效率更高，因此将上述公式左右放大1000倍后可以变换为：

式中括号最后加的500是作四舍五入用。将公式中的浮点数转化为整型后，运算效率得到一定的提升，唯一耗时的部分在于最后的除法运算，因此可以考虑将除法运算替换成最快的位移运算。采用16位精度来计算R、G、B前面的系数：

采用去尾法得到一个整数的系数，最后得到的灰度转换公式为：

实验结果：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 公式 | 视频长度（帧） | 视频分辨率 | 总时间（ms） | 平均时间（ms） |
| 1 | 3623 | 1280\*720 | 44651.5 | 12.32 |
| 2 | 12365.4 | 3.41 |
| 3 | 8222.18 | 2.27 |
| Opencv | 11741.9 | 3.24 |



1. **视频中运动物体检测与轮廓提取**

目前检测视频中运动区域比较流行的方法有背景差分法、帧间差分法和光流法。每一种方法都有其各自的优点和局限性，本小节先对光流发进行简单介绍，然后通过实验对比背景差分法和帧差法之间的优劣。

1. **图像灰度化处理**

光流是空间运动物体在观测成像面上的像素运动的瞬时速度。光流法检测运动物体的基本原理是：对平面图像中的每一个像素点赋予一个速度矢量（包含x，y方向的分量），这就形成了一个图像运动的速度矢量场，在运动的一个特定时刻，图像上的点与三维物体上的点一一对应，这种对应关系可由投影关系得到，根据各个像素点的速度矢量特征，可以对图像进行动态分析，检测图像中的运动区域。如果图像中没有运动物体，则光流矢量在整个图像区域是连续变化的。当图像中有运动物体时，目标和图像背景存在相对运动，运动物体所形成的速度矢量必然和邻域背景速度矢量不同，从而检测出运动物体及位置。

光流法大致可以分为三类：基于梯度的方法、基于匹配的方法和基于频域的方法。光流法的优点在于光流不仅有物体的运动信息，还有关于物体三维结构的丰富信息，但是光流算法大多数计算比较耗时，实时性和实用性都较差，较难应用于高清实时监控中，因此，本文不对光流法进行详细讨论分析。

1. **基于背景差分法的运动物体检测**

背景差分法简单的说就是用当前视频帧与背景图像对应的像素点做差从而得到运动物体。用背景差分法进行运动物体检测的性能取决于背景建模的技术，在取得背景图像的情况下，背景差分法速度快，运动物体检测准确，实现简单。但是现实中的监控视频由于环境光变化、场景情况复杂、摄像机抖动等因素影响，使得背景建模通常需要一定的时间，也使得背景建模变得比较困难。

背景建模算法也很多，目前比较常用的背景建模方法有均值法背景建模、中值法背景建模、高斯背景建模等。

1. **中值法背景建模**

中值法背景建模是最简单的背景建模之一，它是一种基于排序的背景建模方法，原理是对连续N帧图像相同位置的像素点灰度值进行从小到大排序，如果数组长度为奇数则此位置像素点的背景灰度值就是有序数组的中间值，如果数组长度为偶数则取数组中间的两个数的平均值作为背景灰度值。即：

式中代表对第k帧到第k+N帧建模后的背景图像，代表在位置为的像素点数组的中间值。此方法虽然简单，但是缺点也是显而易见的，它需要额外的内存来存储N帧图像的灰度值，并且N的取值也直接影响到背景建模的质量，更新背景模型也需要用到二分插入法，造成了额外的时间开销。

实验效果如下图所示：

1. **均值法背景建模**

均值法背景建模也是最简单的背景建模之一，它的原理跟中值法背景建模有点相似，只不过是将中值替换成均值。由经验很容易得出，在固定摄像头拍摄下，运动物体在画面中同一个地方停留的时间很短，例如在视频样本中抽取一帧如图xxx所示，车辆即是运动物体，我们追踪每一帧某一点的像素灰度值变化，可以得到图xxx

从图xxx可以看到，这一点的像素灰度值在通常情况下总是在一个很小的区间范围波动，只有出现运动物体的时候才会有较大的波动，均值法背景建模的原理就是将连续图像序列同一位置的像素灰度值累加后求平均值，并将此平均值作为背景图像的灰度值。即：

使用此方法的建模速度会比较快，因为累加求和可以用于每一帧，而且不需要额外的存储空间来保存图像序列，但是随着时间的增加，这种方法不能很好的适应复杂背景变化，需要重新从0开始建模。

实验效果如下图所示：

1. **高斯背景建模法**

而其中背景建模效果最好，最能适应背景变化的就是高斯背景建模方法。

1. **单高斯背景模型**

单高斯背景模型的中心思想是，认为一个背景图像，其特定位置的像素亮度（灰度值）的分布满足高斯分布，即对背景图像B，坐标为(x,y)的点的灰度值I

满足：

其中u代表平均值，d代表方差，这是背景模型每个像素点的属性。

背景建模过程也就是将图像中每一个像素点进行高斯分布建模，其样本就是视频连续图像中的对应像素点的灰度值。

1. **基于帧差法的运动物体检测**

帧差法是通过对视频图像序列相邻两针或者隔几帧进行一次差分运算，从而获取运动物体区域。与背景差分法相比帧差法少了背景建模的过程，速度是最快的运动物体检测算法，而且帧差法基本不会受到环境光的影响，能够适应各种复杂的环境，稳定性高。但是帧差法也有显而易见的缺点，由于两帧相减，相同的区域差值变为0，因此通过这种方法获得的差图像往往是运动对象的轮廓，不能够提取完整的运动对象，也就是会形成“空洞”，此外差图像中会留下上一帧运动对象的“拖影”。目前也有一些帧差法的改进算法，接下来本文实现并对比了3种帧差算法的实际效果。

1. **传统帧差法**

用相邻两帧直接做差是最简单的方法，其算法步骤如下：

1. 采集视频图像的第k帧和第k-1帧
2. 将和进行灰度化和中值滤波操作，得到和
3. 计算差图像：
4. 对进行阈值化操作，将中灰度值大于阈值T的像素点标记为运动像素点，反之标记为背景，最后得到二值差分图像，如下所示：

实验结果如下图：

1. **三帧差分法**

三帧差分法主要为了解决差图像的“拖影”现象，更精确的得到运动对象的轮廓，与传统帧差法相比多了一步“与”的操作，其算法步骤如下：

1. 采集视频图像序列的连续三帧：第k-1帧，第k帧和第k+1帧
2. 将图像转化为灰度图后进行中值滤波操作得到，和，然后分别计算相邻帧之间的差图像：
3. 对和分别进行二值化操作，得到两个二值和图像和，将这两个图像按像素点进行“与”操作，即可得到第k帧图像的三帧差分结果，公式如下：

实验结果如下图：

1. **与运动历史图结合的三帧差分法**

三帧差分法确实能有效消除“拖影”现象，但是还是没能解决运动物体检测过程中的“空洞”问题。在实际运用过程中，“拖影”的影响只是会造成检测到的矩形轮廓偏大，这对后续的运动物体跟踪没的影响较小，但是“空洞”问题会使得一个运动对象分为两块或者更多，使得运动物体跟踪不准确，影响较大。为了解决“空洞”问题，这里采用了一种与运动历史图像相结合的三帧差分法进行运动区域检测算法。

运动历史图像（MHI）是一种运动分割方法。运动历史图像是一种保存运动物体运动痕迹的数据结构，可以将其视为一个大小与原图像相同的二维数组，在每帧的运动物体检测后对其进行更新，在有运动像素的对应数组坐标以时间的方式做一个标记，随着帧数的推移，不断对运动历史图进行更新，旧的运动像素点如果超过设置的时间区段时就会被清零。

在3.3.2节中得到视频第k帧的三帧差分图像后，采用如下公式对第k帧的运动历史图像进行更新：

其中*t*是运动像素保留时间，单位是帧。当一个像素点经过*t*帧都没有运动时，就清0。*t*的取值如果太大，运动区域检测结果就会造成很长的“拖影”，相反*t*的取值太小，就会形成较大的“空洞”，当*t*=1的时候就跟三帧差分法的结果一样了。

试验结果如下图：

1. **基于**